

## **Formative Assessment: Promises or problems?**

Paul Black  
King's College London

### **A story of success**

Since the group at King's started work on formative assessment in 1996, our two booklets, *Inside the Black Box* (Black & Wiliam 1998a) and *Working Inside the Black Box* (Black et al. 2002) and our book based on the work with schools (Black et al. 2003) have all been 'best sellers', whilst our research article reviewing the evidence that formative reform really raises standards (Black & Wiliam 1998b) has been widely quoted. The ideas and practices have been advertised and used in both the Key Stage 3 strategy and in the Primary National Strategy, and I and many others still receive many invitations to give talks and to run INSET training. But whilst all this sounds like a success story, and I and colleagues have been cheered to see the many productive changes that have come about in schools, we have also met examples where practical implementation seems to be based on limited understanding and superficial adoption. So this article is not a fair review, but rather unfairly negative in focusing on aspects of this work which can inhibit the main potential reward of formative assessment – the improvement of pupils' capacity to learn.

### **Shallow understanding**

It is essential to work to a careful definition of formative assessment:

*An assessment activity can help learning if it provides information to be used as feedback, by teachers, and by their [students] in assessing themselves and each other, to modify the teaching and learning activities in which they are engaged. Such assessment becomes 'formative assessment' when the evidence is actually used to adapt the teaching work to meet learning needs. (Black et al. 2002)*

A frequent misunderstanding is that any assessment by teachers, and in particular the use of a weekly test to produce a record of marks, constitutes formative assessment. It does not. Unless some learning action follows from the outcomes, such practice is merely frequent summative assessment: the key feature, interaction through feedback, is missing. Another misunderstanding is the belief that this is about the coursework assessment that forms part of some GCSEs; such assessment cannot aid learning unless there is active feedback to improve pupils' work as it develops.

### **Much questioning – little dialogue?**

Formative interaction in classrooms involves far more than having good questions: what matters is both the question and the way in which the teacher handles the responses which it elicits. If a question is asked, but pupils given little time to think, with the inevitable one-word responses rewarded if correct and brushed aside if wrong, then there is no formative dialogue as the teacher continues along his/her pre-determined path. Such recitation styles, well described by Alexander (2006, p.14) as 'cognitively restricting rituals' are all too common.

Questions should open up a dialogue; pupils being given time to discuss issues with one another, and then invited to report their ideas to the class. The teacher's role then is to listen, and to respond constructively in the light of what has emerged. What is essential was expressed by a pupil's comment, to a King's researcher, about her teacher:

*Now I know she is interested in what I think, not in whether I've got the right answer*

For formative interaction, wrong, or partly right answers are what the teacher needs to know. The task is to respond to what has been elicited in order to develop discussion amongst the pupils, collecting various ideas, summarizing and then challenging them with further questions that indicate a positive direction for the thinking. The pathology that must be avoided is for the teacher, having heard some responses, to pronounce judgments, positive or negative, and then to carry on as if the responses were irrelevant. The task of promoting learning is far more difficult than this, and yet it is a delicate leadership task, requiring a balance between letting discussion wander at random, and controlling it so tightly that pupils do not feel free to express their ideas.

The crucial indicator here is the quality of the pupils' contributions. Whilst it is natural that, as I have often found, teachers tend to focus on their 'performance' and on that of colleagues when reflecting on or observing classroom events, it is equally important for them to look at the effects on pupils. If pupils are to be helped to progress as learners, they must be led to participation in reasoned discussion, which means that they should be expressing arguments in fully formed sentences using such words as 'think' and 'because', not merely giving answers as short phrases or single words.

One obstacle here is the belief of many teachers that pupils' talk is mainly for social and affective purposes, the real learning being located in writing. The evidence is that a rich diet of spoken language is a powerful aid to learning. Moreover, for many pupils the school may be the only place on which they can experience and develop the habits of reasoned talk see Alexander (2006).

### **It's the comments that matter**

The research showing that a diet of marks does not improve learning, and that comments can do so only if pupils are not distracted by marks that go with them (Butler, 1988) has been a challenge to the common practice in many schools. However, the value lies in the comments, not in the mere absence of marks: one investigation has shown that to simply remove marks with no effort to help with comments is of questionable value (Smith & Gorard, 2005). Teachers who abandon the giving of marks and devote to producing effective comments find that pupils begin to read their comments and to use them to improve their work, thereby ensuring that the effort invested in written work, by both teacher and pupils, serves to support further learning, rather than to merely produce numbers in a record book.

If their comments are to be part of a learning dialogue, they must identify what needs to be done to further improve, and to give the recipient clear guidance on how to improve. As one pupil complained:

*Marking like 'Use paragraphs!' is useless – if I knew how to use them I would have done.*

Concentration solely on comments makes a decisive shift in the character of feedback – away from competitive and summative judgments and towards help in the process of learning. The fundamental point here is that it is wrong to assume that the need to motivate pupils is best served by offering rewards, such as grades or merits. If pupils see a learning exercise as a competition, they foresee that there will be losers as well as winners: those who have a track record as losers will see little point in trying. The aim should be to motivate everyone, even although some are bound to achieve less than others. In pursuing this aim, the type of feedback given is very important.

As Butler's research showed, those given feedback as marks are likely to see it as a way to compare themselves with others (ego-involvement), those given only comments see it as helping them to improve (task-involvement): the latter group out-performed the former. The common practice of giving pupils a continual diet of marks and grades can focus attention on their 'ability' rather than on the importance of effort; as Dweck's research (2000) has shown such a diet damages the self-esteem of low attainers and leads high-attainers to be reluctant to take risks as failure will damage their ego orientation.

Feedback that focuses on what needs to be done can encourage all to believe that they can improve, so that they are motivated to invest effort in the work. Given this, it is regrettable that many teachers who would like to abandon the frequent assignment of marks are prevented from doing so by a school policy which ignores the evidence in requiring that marks be given and reported frequently.

### **A target too far? Peer-assessment and self-assessment**

Pupils can only achieve a learning goal if they understand that goal and can assess what they need to do to reach it – self-assessment is essential to learning (Sadler, 1989). In practice, peer-assessment turns out to be an important complement to self-assessment. Peer work is particularly valuable because the interchange will be in language that pupils themselves would naturally use, and because pupils learn by trying to teach, and by being assessed by, their peers (Sadler, 1998).

However, pupils are often seen working *in* groups but not working *as* groups. Dawes et al. (2004) have explored the training that pupils need in order to work creatively together, and his rules, that all must participate, that all contributions must be treated with respect, that a group must achieve consensus, and above all, that all claims must be supported by reasons, have been shown to lead to improved attainment. A meta-analysis of numerous studies (Johnson et al. 2000) has shown that groups in which learners collaborate yield big learning gains over either individual study or work where there is competition within each group. Groups where there is such internal competition produce almost no learning advantage over individual learning. Thus whilst group work is very common in classrooms, it is its quality, and not its mere existence, which determines the help it will give to pupils' learning.

In developing peer- and self-assessment skills, many find that the most difficult task is help

pupils to appraise their work in relation to learning targets. Whilst pupils can only steer their own work if they can see it in the light of a learning target, they must understand what the target means, and have some idea of what it would be like to attain it – i.e they must grasp the criteria of quality which they must apply in auditing any work. The delicate task is then to choose and specify targets, and the accompanying criteria, which are close enough to pupils' understanding that they are meaningful to them, and far enough ahead that they present a realistic challenge. Understanding criteria is often the hardest part - it can be helped if pupils engage in modeling exercises, or in exercises where they rank order pieces of work of varying quality and then discuss how such judgments are justified.

National Curriculum levels, designed so that the interval between them represents two years of learning progress, are quite inadequately coarse for this purpose<sup>1</sup>. Labeling pupils according to level expectations or to GCSE grades, and to frequently repeat or up-date these, will not itself help learning progress, and may, by producing the negative effects of ego-involvement, be self-confirming where pupils conclude that they may only achieve low-grade results. Such practices are the learning equivalent of pulling up the growing plant in order to check its roots. A summative overview guide is of course essential from time to time, but not so frequently that it dominates and detracts from the day-to-day process of learning. The fact that parents and pupils want frequent marks and level predictions may be an appetite created, in part, by schools themselves, who might do better to help all partners to focus more on the detailed steps needed for the improvement of learning.

### **Expecting too much - too quickly?**

The overall message here is that there are many ways in which 'doing assessment for learning' may fail to implement those key features whereby pupils are helped to become more confident and effective learners, and thereby achieve higher standards of attainment. There are two main reasons for this.

Many of the most important changes involved make heavy demands on teachers and schools in that they require them to re-think their roles in their core task of promoting pupils' learning. In the work of the King's group with a set of six schools, it took two years before changes became embedded in the teachers' classroom practices – there was little sign of change after only one year. Yet this was in a project where all the teachers had a whole-day meeting together at King's once every five weeks, research staff visited each individually to observe lessons and give feedback, and they were keen to adopt the innovations.

A second reason is the failure to understand the support that teachers must have if they are to make radical changes. Inspirational talks, or brief training courses, or reams of written advice, may be useful to inspire and guide, but given the personally challenging nature of what is called for, it should not be surprising that such methods are, on their own, inadequate. The essential requirement is that teachers should be supported by a commitment sustained within their whole school over several years, a commitment which arranges that they have time to share experiences and advice with colleagues, and that they can risk, and be helped to overcome and learn from, failures. Moreover, their schools should not choose, and should not be required, to commit their staff to other innovations until this one is well established.

## References

- Alexander, R. (2006) *Towards Dialogic teaching: rethinking classroom talk*. 3rd Edn. Cambridge UK: Dialogos
- Black, P. & Wiliam, and D. (1998a) *Inside the Black Box: raising standards through classroom assessment* London UK: nferNelson.
- Black, P. & Wiliam, D. (1998b) Assessment and Classroom Learning. *Assessment in Education*. **5** (1) 7-74
- Black, P. J.; Harrison, C.; Lee, C.; Marshall, B. & Wiliam, D (2002) *Working inside the black box: Assessment for learning in the classroom*. London, UK: nferNelson
- Black, P.; Harrison, C.; Lee, C.; Marshall, B. & Wiliam, D. (2003). *Assessment for learning: putting it into practice*. Buckingham, UK: Open University Press.
- Butler, R. (1988). Enhancing and undermining intrinsic motivation; the effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, **58**, 1-14.
- Dweck, C. S. (2000), *Self-Theories: Their role in motivation, personality and development*, London UK: Taylor & Francis
- Johnson, D.W., Johnson, R.T. & Stanne, M.B. 2000 Co-operative learning methods: a meta-analysis. Down-loadable from: <http://www.co-operation.org/pages/cl-methods.html>
- Dawes, L., Mercer, N. & Wegerif, R. (2004) *Thinking Together*. 2<sup>nd</sup> edn. York UK: Imaginative Minds
- DES (1988) *Task Group on Assessment and Testing - National Curriculum: A Report*. London: Department of Education and Science.
- Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, **18**, 119-144.
- Sadler, R. (1998). Formative assessment: revisiting the territory. *Assessment in Education* **5**(1), 77-84
- Smith, E & Gorard, S. (2005) 'They don't give us our marks': the role of formative feedback in student progress. *Assessment in Education* **12**(1) 21-38.
- Wiliam, D.; Lee, C.; Harrison, C. & Black, P. (2003). Teachers developing assessment for learning: impact on student achievement. *Assessment in Education*, **11**(1), 49-65

---

<sup>i</sup> The idea of levels, originally proposed in the TGAT report (DfES 1988), was that they would be built in to the structure of each curriculum so that it would set out the expected progress of pupils; thus, for example, the components of level 2 would be the average of pupils' achievement at age 7, level 4 at age 11 and so on. This is misinterpreted when level 4 is now regarded as the target for *all* to achieve at age 11: the point was to allow for the well-established facts, that there is a range in pupils' achievements at any one age which can be at least as wide as the progress of the average pupil over a few years, and that this range increases with age.